UCL
Biochemistry

# DEPARTMENT OF BIOCHEMISTRY & MOLECULAR BIOLOGY

About the Department
Research Activities
PhD Studentships
Undergraduate Admissions
Department Staff
Seminars
Undergraduate Teaching Resources
Job Vacancies
Location & Map

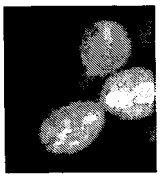
## Welcome to the Department of Biochemistry & Molecular Biology...

We aim to provide a stimulating and diverse research and training environment of international standing within which important and exciting areas of modern biology, biotechnology and medicine can be investigated at the atomic, molecular, cellular and organism level.

The department pursues these research aims through its high level of external research funding, its investment in excellent new facilities, up-to-date equipment and state-of-the-art technologies, the recruitment of high-calibre staff, maintenance of industrial contacts, and the fostering of close links and collaborations within the extensive UCL and Bloomsbury scientific community.

The disciplines of biochemistry and molecular biology have never been more relevant to the furtherance of our fundamental knowledge and ability to exploit biological systems than in this, the 'post-genomic' era. From the structural biology of proteins and enzymes to the mechanism of amphibian limb regeneration; from the regulation of transcription of genes involved in drug metabolism to the uncovering of gene function in Mycobacterium tuberculosis, from understanding the signalling of insulin receptors to the computer analysis of whole genomes sequences, this department provides an exciting venue in which to realise the promise of these important goals.

And to assist us in this endeavour, during the period 2003-2005 the department will enjoy an investment of around £9 million in renewed infrastructure and research facilities.



>> New Website message from th Webmaster...

>> Events

Department of Biochemistry & Molecular Biology - University College London - Gower Street - London - WC1E 6BT

☎ +44 (0)20 7679 2329/7033 Fax +44 (0)20 7679 7193

This page was last modified October 6, 2003 by the Webmaster

# Antibodies - General Information

This page summarises a lot of generally useful information about antibodies.

## The Kabat Numbering Scheme

The Kabat numbering scheme is a widely adopted standard for numbering the residues in an antibody in a consistent manner. However the scheme does have its problems!

First, since the numbering scheme was developed from (fairly limited) sequence data, the position at which insertions occur in CDR-L1 and CDR-H1 does not match the structural insertion position. Thus topologically equivalent residues in these loops do not get the same number.

Second, the numbering adopts a rigid specification. For example in the potentially very long CDR-H3, insertions are numbered between residue H100 and H101 with letters up to K (i.e. H100, H100A ... H100K, H101). If there are more residues than that, there is no standard way of numbering them. Such situations occur at other positions too.

The numbering throughout the chains follows.

### Light chain

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | | |
| 27A | 27B | 27C | 27D | 27E | 27F | | | 28 | 29 |
| 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 |
| 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 |
| 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 |
| 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 |
| 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 |
| 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 |
| 90 | 91 | 92 | 93 | 94 | 95 | | | | |
| 95A | 95B | 95C | 95D | 95E | 95F | 96 | 97 | 98 | 99 |
| 100 | 101 | 102 | 103 | 104 | 105 | 106 | | | |
| 106A | | | | | | | 107 | 108 | 109 |

### Heavy chain

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| 30 | 31 | 32 | 33 | 34 | 35 | | | | |
| 35A | 35B | | | | | 36 | 37 | 38 | 39 |
| 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 |
| 50 | 51 | 52 | | | | | | | |
| 52A | 52B | 52C | 53 | 54 | 55 | 56 | 57 | 58 | 59 |
| 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 |
| 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 |
| 80 | 81 | 82 | | | | | | | |
| 82A | 82B | 82C | 83 | 84 | 85 | 86 | 87 | 88 | 89 |
| 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |
| 100 | | | | | | | | | |
| 100A | 100B | 100C | 100D | 100E | 100F | 100G | 100H | 100I | 100J |
| 100K | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 |

```
110   111   112   113
```

# The Chothia Numbering Scheme

The Chothia numbering scheme is identical to the Kabat scheme, but places the insertions in CDR-L1 and CDR-H1 at the structurally correct positions. This means that topologically equivalent residues in these loops do get the same label (unlike the Kabat scheme). There are two disadvantages: first, the Kabat scheme is so widely used that some confusion can arise; second, Chothia et al. **changed** their numbering scheme as of their 1989 Nature paper such that insertions in CDR-L1 are placed after residue L31 rather than L30. Examining the conformations of the loops shows that **L30 is the correct position**.

The pre-1989 Chothia numbering (the structurally correct version) throughout the chains follows.

**Note** That in their latest paper (Al-Lazikani et al., (1997) JMB **273**,927-948), Chothia's group returns to using residue 30 as the insertion site in CDR-L1**!**

## Light chain

```
 0     1     2     3     4     5     6     7     8     9
10    11    12    13    14    15    16    17    18    19
20    21    22    23    24    25    26    27    28    29
30
30A   30B   30C   30D   30E   30F
      31    32    33    34    35    36    37    38    39
40    41    42    43    44    45    46    47    48    49
50    51    52    53    54    55    56    57    58    59
60    61    62    63    64    65    66    67    68    69
70    71    72    73    74    75    76    77    78    79
80    81    82    83    84    85    86    87    88    89
90    91    92    93    94    95
95A   95B   95C   95D   95E   95F   96    97    98    99
100   101   102   103   104   105   106
106A                                      107   108   109
```

## Heavy chain

```
 0     1     2     3     4     5     6     7     8     9
10    11    12    13    14    15    16    17    18    19
20    21    22    23    24    25    26    27    28    29
30    31
31A   31B
            32    33    34    35    36    37    38    39
40    41    42    43    44    45    46    47    48    49
50    51    52
52A   52B   52C   53    54    55    56    57    58    59
60    61    62    63    64    65    66    67    68    69
70    71    72    73    74    75    76    77    78    79
80    81    82
82A   82B   82C   83    84    85    86    87    88    89
90    91    92    93    94    95    96    97    98    99
100
100A  100B  100C  100D  100E  100F  100G  100H  100I  100J
100K  101   102   103   104   105   106   107   108   109
110   111   112   113
```

# Table of CDR Definitions

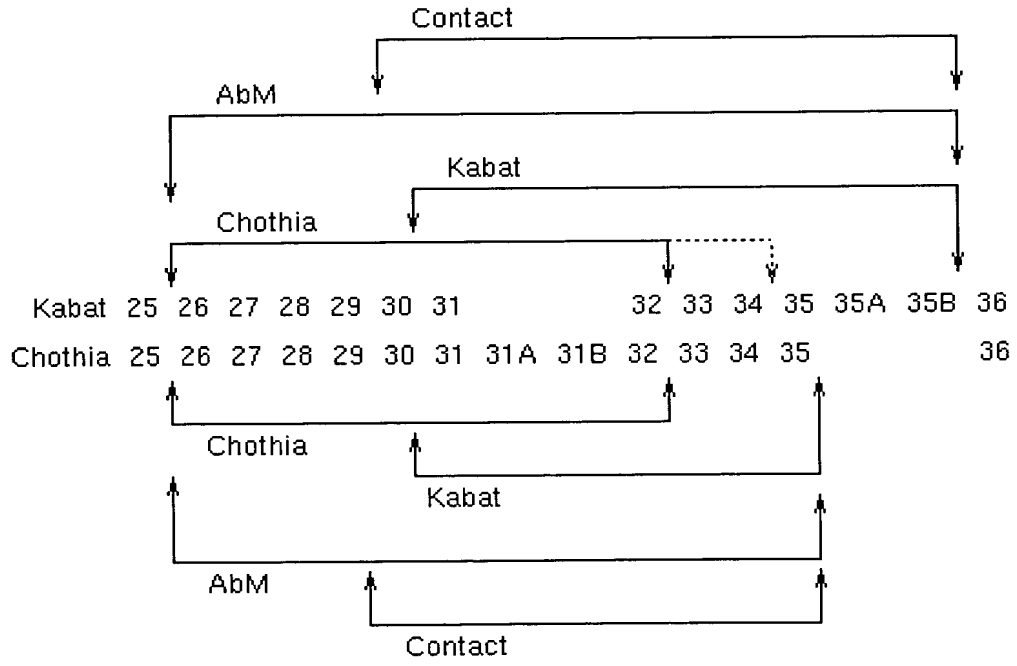
A number if definitions of the CDRs are commonly in use:

- The Kabat definition is based on sequence variability and is the most commonly used.
- The Chothia definition is based on the location of the structural loop regions.
- The AbM definition is a compromise between the two used by Oxford Molecular's AbM antibody modelling software.
- The contact definition has been recently introduced by us and is based on an analysis of the available complex crystal structures. This definition is likely to be the most useful for people wishing to perform mutagenesis to modify the afinity of an antibody since these are residues which take part in interactions with antigen.

| Loop | Kabat | AbM | Chothia | Contact |
|---|---|---|---|---|
| L1 | L24 -- L34 | L24 -- L34 | L24 -- L34 | L30 -- L36 |
| L2 | L50 -- L56 | L50 -- L56 | L50 -- L56 | L46 -- L55 |
| L3 | L89 -- L97 | L89 -- L97 | L89 -- L97 | L89 -- L96 |
| H1 | H31 -- H35B | H26 -- H35B | H26 -- H32..34 | H30 -- H35B |
| | (Kabat Numbering) | | | |
| H1 | H31 -- H35 | H26 -- H35 | H26 -- H32 | H30 -- H35 |
| | (Chothia Numbering) | | | |
| H2 | H50 -- H65 | H50 -- H58 | H52 -- H56 | H47 -- H58 |
| H3 | H95 -- H102 | H95 -- H102 | H95 -- H102 | H93 -- H101 |

Note that the end of the Chothia CDR-H1 loop when numbered using the Kabat numbering convention varies between H32 and H34 depending on the length of the loop. (This is because the Kabat numbering scheme places the insertions at H35A and H35B.)

- If neither 35A nor 35B is present, the loop ends at 32
- If only 35A is present, the loop ends at 33
- If both 35A and 35B are present, the loop ends at 34



This diagram illustrates the alternative definitions for CDR-H1. The Kabat and Chothia numbering schemes are shown horizontally and the Kabat, Chothia, AbM and Contact definitions of the CDRs are shown with arrows above and below the two numbering schemes.

# Table of mean contact data

Following an analysis of the contacts between antibody and antigen in the complex structures available in the Protein Databank, we have generated a set of mean contact data. The full method by which these results were obtained is described in the following paper: MacCallum, R. M., Martin, A. C. R. and Thornton, J. T. **Antibody-antigen interactions: Contact analysis and binding site topography.** J. Mol. Biol. **262,** 732-745.

Briefly, we have analysed the number of contacts made at each position, defining contact as burial by > 1 square Angstrom change in solvent accessibility. These data give a simple measure of how likely a residue is to be involved in antigen contact.

Second, we have calculated the mean percentage burial over the accessible residues.

Click here for an image showing a composite combining site containing all CDR conformations coloured by contact propensity.

The table presents the chain name, residue number (N.B. This is **pre-1989 Chothia Numbering**), the number of contacts and the mean percent burial. The data are available by clicking here.

An alternative simplified view is presented as a list of CDR residues making contact in each antibody with summary data for each CDR.

# How to identify the CDRs by looking at a sequence

## CDR-L1

Start - Approx residue 24
Residue before is **always** a Cys
Residue after is **always** a Trp. Typically TRP-TYR-GLN, but also, TRP-LEU-GLN, TRP-PHE-GLN, TRP-TYR-LEU
Length 10 to 17 residues

## CDR-L2

Start - **always** 16 residues after the end of L1
Residues before generally ILE-TYR, but also, VAL-TYR, ILE-LYS, ILE-PHE
Length **always** 7 residues (except 7FAB which has a deletion in this region)

## CDR-L3

Start - **always** 33 residues after end of L2 (except 7FAB which has the deletion at the end of CDR-L2)
Residue before is **always** Cys
Residues after **always** PHE-GLY-XXX-GLY
Length 7 to 11 residues

## CDR-H1

Start - Approx residue 26 (**always** 4 after a CYS) [Chothia / AbM defintion] Kabat definition starts 5 residues later
Residues before **always** CYS-XXX-XXX-XXX
Residues after **always** a TRP. Typically TRP-VAL, but also, TRP-ILE, TRP-ALA
Length 10 to 12 residues (AbM definition) Chothia definition excludes the last 4 residues

## CDR-H2

Start - **always** 15 residues after the end of Kabat / AbM definition) of CDR-H1
Residues before typically LEU-GLU-TRP-ILE-GLY, but a number of variations
Residues after LYS/ARG-LEU/ILE/VAL/PHE/THR/ALA-THR/SER/ILE/ALA
Length Kabat definition 16 to 19 residues (AbM definition ends 7 residues earlier)

## CDR-H3

Start - **always** 33 residues after end of CDR-H2 (**always** 2 after a CYS)
Residues before **always** CYS-XXX-XXX (typically CYS-ALA-ARG)
Residues after **always** TRP-GLY-XXX-GLY
Length 3 to 25(!) residues

214–218 Nucleic Acids Research, 2000, Vol. 28, No. 1 

# Kabat Database and its applications: 30 years after the first variability plot

**George Johnson and Tai Te Wu***

Departments of Biochemistry, Molecular Biology and Cell Biology, and of Biomedical Engineering, Northwestern University, Evanston, IL 60208, USA

Received July 23, 1999; Revised and Accepted October 13, 1999

## ABSTRACT

**The Kabat Database was initially started in 1970 to determine the combining site of antibodies based on the available amino acid sequences at that time. Bence Jones proteins, mostly from human, were aligned, using the now-known Kabat numbering system, and a quantitative measure, variability, was calculated for every position. Three peaks, at positions 24–34, 50–56 and 89–97, were identified and proposed to form the complementarity determining regions (CDR) of light chains. Subsequently, antibody heavy chain amino acid sequences were also aligned using a different numbering system, since the locations of their CDRs (31–35B, 50–65 and 95–102) are different from those of the light chains. CDRL1 starts right after the first invariant Cys 23 of light chains, while CDRH1 is eight amino acid residues away from the first invariant Cys 22 of heavy chains. During the past 30 years, the Kabat database has grown to include nucleotide sequences, sequences of T cell receptors for antigens (TCR), major histocompatibility complex (MHC) class I and II molecules and other proteins of immunological interest. It has been used extensively by immunologists to derive useful structural and functional information from the primary sequences of these proteins. An overall view of the Kabat Database and its various applications are summarized here. The Kabat Database is freely available at http://immuno.bme.nwu.edu**

## INTRODUCTION

The purpose of maintaining the Kabat Database of aligned sequences of proteins of immunological interest, in our opinion, is to provide useful correlations between structure and function for this special group of proteins from their nucleotide and amino acid sequences to their tertiary structures (1). These sequences are thus aligned with the ultimate aim of understanding how these proteins are folded and how they can perform their biological functions. We include only coding region sequences that have been published. In some cases, only the amino acid sequences were published, while the corresponding nucleotide sequences were deposited in GenBank. All stored sequences were then printed out and checked visually against available published sequences. We routinely survey for possible new sequences in journals in our libraries, Medline entries, cross-references from other papers, and author notification; however, we may still miss some sequences. GenBank, on the other hand, contains a substantial number of unpublished sequences. If there are doubts about these sequences or their annotations, please refer to the original papers. The Kabat numbering systems (see the Introduction of 2) for antibody light and heavy chains, for TCR alpha and beta chains, etc., go hand-in-hand with variability calculations. The locations of the CDRs are the theoretically derived positions which can be verified experimentally. Indeed, from the first antigen–antibody Fab complex (3) to the complexes of TCR, processed peptide and MHC class I molecule (4,5), it has been realized that alignment of amino acid sequences and variability calculations can be of utmost importance in understanding how these important macromolecules function biologically. Due to the rapid development of genetic and protein engineering methods, mouse and rat antibodies have been humanized to treat human cancers, viral infections, etc (6). CDRs of selected rodent antibodies are cut out and glued onto human antibody frameworks to minimize rejection by human patients.

Our predicted CDRs are slightly different from Chothia's. A careful comparison can be found from a hyperlink on our website to 'Andrew's Antibody Page' (http://www.biochem.ucl.ac.uk/~martin/abs/index.html ).

Massive amounts of sequence data are being continuously published in the scientific literature. It is imperative to collect and properly align the sequences so that they can be used by as many researchers in this field as possible. We have previously published five editions of these sequences (see the Introduction of 2). In 1991, the fifth edition (2) consisted of three volumes. Currently, the database is more than five times as large. As of September 29, 1999, the Kabat database contained 1 599 375 and 2 517 756 nt for antibody light and heavy chain variable regions, respectively, as compared to 272 244 and 418 962 nt in 1991. Total numbers of entries, amino acids and bases of other categories of sequences can be obtained by using the 'Current Counts' hyperlink on our website. The collection is available on our website (http://www.immuno.bme.nwu.edu ) which is free due to the generous support by various research grants from NIH since 1970.

Finally, numerous scientific papers have cited our database, quoting our fourth edition (7), fifth edition (2), or one of our more recent papers (8). On our part, we have been analyzing

*To whom correspondence should be addressed. Tel: +1 847 491 7849; Fax: +1 847 491 4928; Email: t-wu@nwu.edu

the Kabat Database during the past few years with reference to the total numbers of antibody and TCR V-genes, possible evolutionary selection processes, importance of antibody CDRH3s as related to their fine specificities, etc.

## KABAT DATABASE

The Kabat Database may be accessed for searching, sequence retrieval and analysis by a few different methods: electronic mail, WWW and ftp. The electronic mail interface has been available since 1993, the WWW interface since 1995 and various formats of the database in electronic format for nearly a decade (8). Our data formats, searching tools, output formats and database structures have gradually been adopted by other immunological databases and interfaces.

### Electronic mail interface

An electronic mail interface (seqhunt2@immuno.bme.nwu.edu ) provides a non-interactive method for searching and sequence retrieval (9). Sending mail to the server address with the single word 'help' (no quotes) in the message body returns instructions for using the server.

All sequences classes are searchable and returnable. The query format allows making AND/OR/NOT constructed restrictions on the database and amino acid and nucleotide sequence pattern matching with allowable differences. Requests are processed as they are received and depending on the network traffic, take ~1–2 min to be searched and returned to the sender. The returned format is a fixed-line length record of 80 or fewer characters per line for ease in visual inspection and processing by user-written scripts or programs. The characters are plain text.

The query format for the sent request consists of two parts. The first part contains directives for the server to follow while the second part contains specifications of the search. Specification of the extent of data returned, the number of documents to return, starting document and whether plain ASCII text or PostScript should be used in the return format may be entered. Further, one can direct the server to return a distribution, the variability or unaligned raw data for the search specified.

The second part of the query contains the search restrictions on the database. Words separated by AND and OR may be used, as well as searching functions, like nucleotide/amino acid pattern matching and positional restriction matching.

There are basically three steps in translating and performing a search on the Kabat Database: generate the question or query, translate it into a format the server can recognize and decide on the output options desired of the returned matches. For example, if matches of mouse kappa light chains of anti-phosphorylcholine antibodies are desired, the query and restriction on the database would be:

Begin
@mouse and kappa and phosphorylcholine

The '@' before mouse tells the server that matches of the species mouse are desired, rather than searching through the entire database record for instances of the word 'mouse'. More complicated restrictions can be generated using parentheses for grouping and the minus sign '–' for NOT. Finding all rat and rabbit sequences which are not kappa light chains, and returning them as amino acid sequences in PostScript format would be constructed as:

PSAA
Begin
(rat and rabbit) and –kappa

Pattern matching is interpreted as the second part of an AND statement, such that finding all rat and rabbit sequences which are not kappa and contain the nucleotide pattern cagtacgtcag with three allowable mismatches, would be sent as:

Begin
(rat and rabbit) and –kappa [ implicit AND ]
#NM 3
cagtacgtcag

More examples of searching and output options may be found in the 'help' file returned from the server.

### WWW interface

The WWW interface (8) to the Kabat Database: http://immuno.bme.nwu.edu contains searching and analysis tools as well as links to database download sites and other interesting databases. Most of the features found in the electronic mail interface are found in the WWW interface, as well as other tools. The WWW interface is more interactive than the Email and returns results faster, depending on the network traffic.

### Searching and analysis tools

*SeqhuntII.* This grouping of programs allows searches through the annotations and sequence pattern matching of the amino acid and nucleotide sequence data with allowable mismatches. Like the Email server, restrictions on the database may be formulated as AND/OR/NOT constructs. Output extent, output format, maximum documents and starting document may be specified. Browsing of the output results in HTML format allows the user to view the database entries in an easy-to-read format. ASCII text may be selected as output for use in user-generated scripts and programs. PostScript generation allows for printing on a PostScript supporting printer. Sequence matching is returned aligned with the target sequence with nucleotide or amino acid differences from the database sequence displayed in a case change. Since the database contains only coding regions of genes and proteins, the query sequence should be a portion of the coding region being sought.

*Variability.* Variability and amino acid distributions of sequence groups may be generated for restrictions on the database. The variability plots are in PostScript format and may either be viewed on the screen with an appropriate PostScript viewer (e.g. GNU ghostscript or ghostview) or printed to a postscript-supporting printer. Plots for human and mouse TCR gamma and delta chain variable regions are shown in Figure 1. Scaling of the variability plots may be done allowing comparison of variability plots for different groupings of sequences. Distributions of the amino acids per position may be returned also, including the calculated variability for each position.

*Sequence alignment.* Alignment of user-entered coding regions of immunoglobulin light chains according to the Kabat numbering system can be performed. Because of the relatively few alignment options available for light chains, most sequences can be aligned. One can start with around 10 amino acid residues or 30 nt. There is no lower limit on the length of sequence to be matched. In some cases though, visual inspection and alignment is necessary, as is for heavy chain alignment,

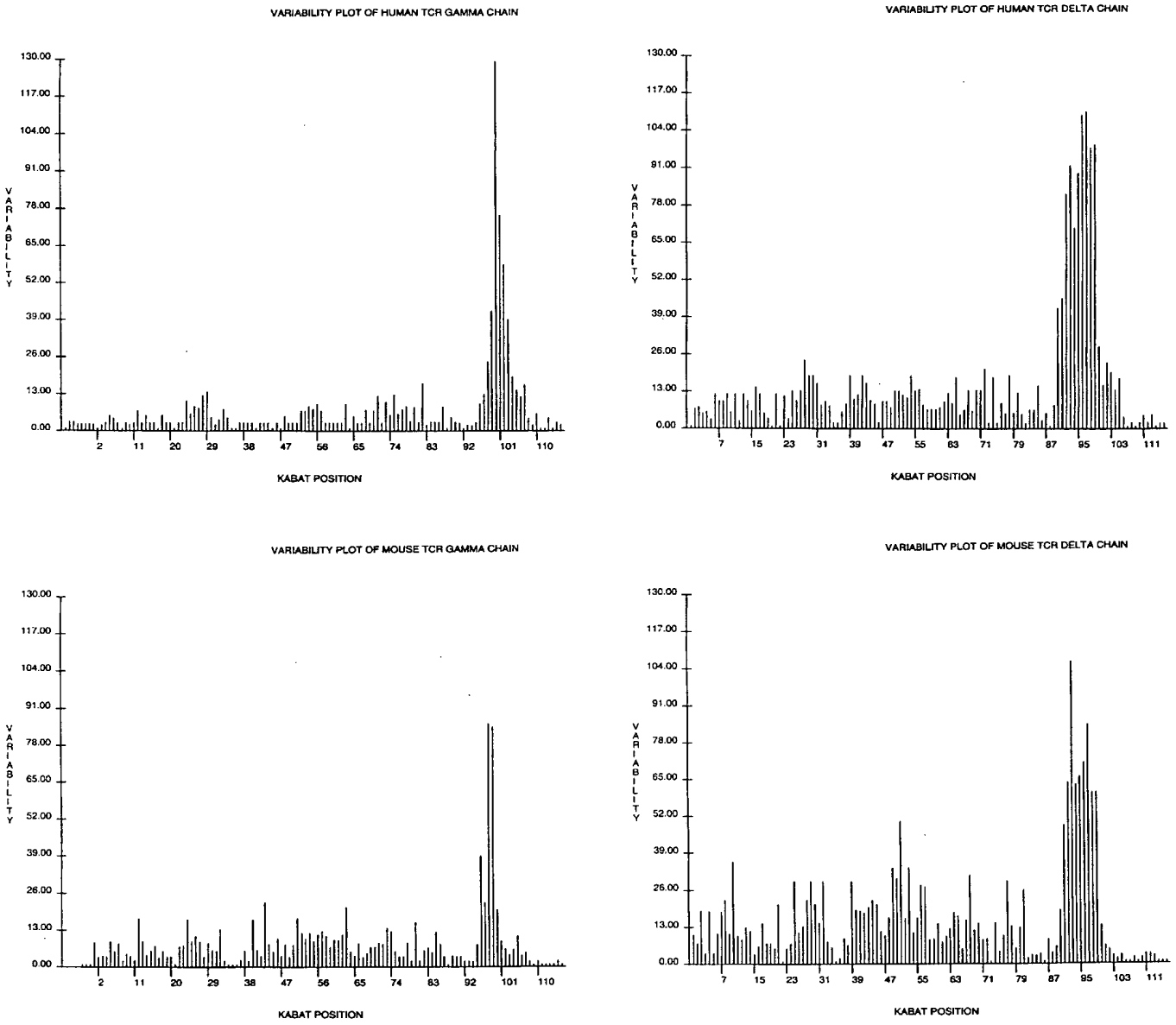



**Figure 1.** Variability plots for human and mouse TCR gamma and delta chain variable regions, using 377 human gamma, 1260 human delta, 297 mouse gamma and 461 mouse delta partial and complete sequences.

especially within the CDRH3 region, if additional codons or residues are inserted and denoted by '#'. If a suitable alignment counterpart from the database is not found for the target sequence, the user can contact us.

*FTP.* Various formats of the database are available for download from NCBI's repository under the directory 'kabat'. Currently active formats include a FASTA-like raw sequence format and the database's fixed length format of 80 or fewer characters per line and vertical alignment. Four main variations on the fixed length format exist to properly visually display single translations, pseudogene translations, J-minigenes and D-minigenes. Other immunological databases have adopted similar formats as exemplified by the three letter code amino acid translation followed by single letter code. A 'dump' version of the database is periodically updated which contains unlimited line length records more suitable for mass processing on unix-based systems.

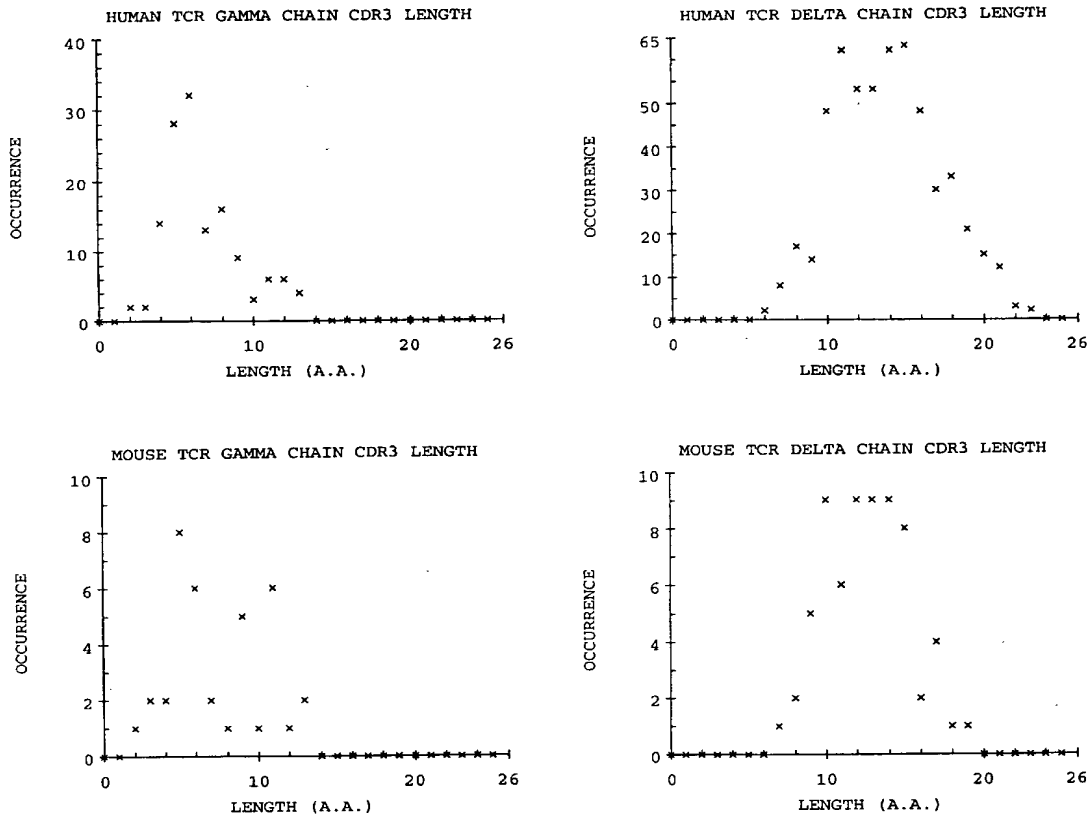



**Figure 2.** Length distributions of CDR3s of human and mouse TCR gamma and delta chains, based on 135 human gamma, 546 human delta, 37 mouse gamma and 66 mouse delta complete CDR3 sequences.

## OTHER APPLICATIONS

As mentioned before, the Kabat Database was initially constructed for the purpose of identifying the antibody combining site (1). Starting from aligned amino acid sequences and using variability calculations, we have identified CDRs of antibody light and heavy chains, as well as those of TCRs. Such calculations can also provide useful predictions for MHC class I and II sequences (8), and to other aligned proteins sequences, e.g. HIV gp120, gp41, etc.

The importance of CDRH3 to confer fine specificity to antibodies was realized a few years ago (10). Furthermore, the unique CDRH3 nucleotide sequences have recently been used as a sensitive diagnostic test to detect residue B cell malignancies in cancer patients. Thus, many of these sequences have been determined. But most of them have been excluded from GenBank due to their relative short lengths. We have been meticulously collecting them, and realized the importance of their length distributions in antibodies of various specificities (11), and possible differences between CDRH3s of human and mouse (12). In the case of rabbit, the CDRH3s have less length variation than human and mouse. This may be compensated by the length variations of the CDRL3s (13).

The length variations of TCR alpha and beta chain CDR3s are very restricted (14). On the other hand, TCR gamma and delta chain CDR3s have more length variation, close to those of antibody heavy chains (Fig. 2). Whether they bind antigens directly is unclear.

During recent years, various research groups have decided to sequence the entire coding region of different antibody and TCR V-genes, in order to have an idea of their total numbers. On the other hand, we have discovered that pair-wise comparisons of V-gene nucleotide sequences in the Kabat Database provide very accurate estimations of their total numbers (15,16). In addition, such comparisons seem to suggest that antibody and TCR V-genes have evolved under different selective pressures (17). In the case of MHC class I sequences, comparison of their aligned sequences has elucidated a new mechanism of generating novel MHC class I molecules by random assortment of their a1 and a2 gene segments (18).

## DISCUSSION

The Kabat Database has been around for 30 years. It has provided the immunology community a useful service, since it

not only is a sequence database but also incorporates vital aspects of the biology of the immune system. Various analytical methods have been developed to study the structure and function relations of proteins of immunological interest.
Electronic addresses:
http://immuno.bme.nwu.edu
seqhunt2@immuno.bme.nwu.edu
Citing the Kabat Database:
Authors using this database may cite this paper together with the electronic addresses.

## ACKNOWLEDGEMENT

Supported in part by NIH Grant 5 R24 AI25616-10.

## REFERENCES

1. Wu,T.T. and Kabat,E.A. (1970) *J. Exp. Med.*, **132**, 211–250.
2. Kabat,E.A., Wu,T.T., Perry,H., Gottesman,K. and Foeller,C. (1991) *Sequences of Proteins of Immunological Interest*, Fifth Edition. NIH Publication No. 91-3242.
3. Amit,A.G., Mariuzza,R.A., Phillips,S.E.V. and Poljak,R.J. (1986) *Science*, **233**, 747–753.
4. Garcia,K.C., Degano,M., Stanfield,R.L., Brunmark,A., Jackson,M.R., Peterson,P.A., Teyton,L. and Wilson,I.A. (1996) *Science*, **274**, 209–219.
5. Garboczi,D.H., Ghosh,P., Utz,U., Fan,Q.R., Biddison,W.E. and Wiley,D.C. (1996) *Nature*, **384**, 134–141.
6. Jones,P.T., Dear,P.H., Foote,J., Neuberger,M.S. and Winter,G. (1986) *Nature*, **321**, 522–525.
7. Kabat,E.A., Wu,T.T., Reid-Miller,M., Perry,H. and Gottesman,K. (1987) *Sequences of Proteins of Immunological Interest*, Fourth Edition. US Govt. Printing Off. No. 165-492.
8. Johnson,G., Kabat,E.A. and Wu,T.T. (1996) In Herzenberg,L.A., Weir,W.M., Herzenberg,L.A. and Blackwell,C. (eds), *Weir's Handbook of Experimental Immunology I. Immunochemistry and Molecular Immunology*, Fifth Edition. Blackwell Science Inc., Cambridge, MA, pp. 6.1–6.21.
9. Johnson,G., Wu,T.T. and Kabat,E.A. (1995) In Paul,S. (ed.), *Antibody Engineering Protocols*. Humana Press, pp.1–15.
10. Kabat,E.A. and Wu,T.T. (1991) *J. Immunol.*, **147**, 1709–1719.
11. Johnson,G. and Wu,T.T. (1998) *Int. Immunol.*, **10**, 1801–1805.
12. Wu,T.T., Johnson,G. and Kabat,E.A. (1993) *Proteins*, **16**, 1–7.
13. Sehgal,D., Johnson,G., Wu,T.T. and Mage,R.G. (1999) *Immunogenetics*, **50**, 31–42.
14. Johnson,G. and Wu,T.T. (1999) *Immunol. Cell Biol.*, **77**, 391–394.
15. Johnson,G. and Wu,T.T. (1997) *Genetics*, **145**, 777–786.
16. Johnson,G. and Wu,T.T. (1997) *Immunol. Cell Biol.*,**75**, 580–583.
17. Johnson,G. and Wu,T.T. (1997) *J. Mol. Evol.*, **44**, 253–257.
18. Johnson,G. and Wu,T.T. (1998) *Genetics*, **149**, 1063–1067.